# Otimização para Motores Generativos em Português Brasileiro: uma replicação do estudo GEO de Princeton em consultas de finanças, saúde e turismo

André Henrique de Paula

*Pesquisador Independente — São Paulo, SP, Brasil — andrehp.eth@gmail.com — ORCID: 0009-0009-1875-1631*

## Resumo

Motores generativos — sistemas que respondem a consultas sintetizando e citando múltiplas fontes recuperadas da web por meio de grandes modelos de linguagem (LLMs) — estão substituindo progressivamente a busca tradicional, e o estudo fundador de Aggarwal et al. (KDD 2024) demonstrou, em inglês, que técnicas de Otimização para Motores Generativos (GEO) podem elevar a visibilidade de uma página em até 40%. Este trabalho apresenta a primeira replicação sistemática desse experimento em português brasileiro. Construímos um banco de 200 consultas reais nos domínios de finanças, saúde e turismo, simulamos um motor generativo de duas etapas (recuperação das 5 primeiras páginas + síntese com citações por LLM) e aplicamos os nove métodos de otimização do estudo original a uma página-alvo, com 3 repetições por condição (N = 6.000 execuções). Os resultados replicam o padrão central de Princeton: adição de estatísticas (+34,2% em PAWC; +28,4% na taxa de citação), adição de citações diretas (+26,0%) e citação de fontes (+22,3%) produzem ganhos significativos (Wilcoxon pareado, $p < 0,01$), enquanto keyword stuffing — a tática clássica de SEO — reduz a visibilidade em 25,2%. Observamos ainda o "efeito equalizador": páginas mal ranqueadas no Google (posições 4–5) obtêm ganhos proporcionais de citação (+33% a +44%) muito superiores aos das primeiras posições (+19%). Os efeitos são heterogêneos por domínio, com saúde exibindo o padrão mais conservador. Os achados indicam que os princípios de GEO se transferem para o ecossistema lusófono e oferecem diretrizes empíricas para criadores de conteúdo brasileiros.

**Palavras-chave:** otimização para motores generativos; GEO; busca generativa; grandes modelos de linguagem; visibilidade de conteúdo; português brasileiro.

## Abstract

Generative engines answer user queries by synthesizing and citing multiple retrieved web sources through large language models, and the founding GEO study (Aggarwal et al., KDD 2024) showed, in English, that content optimization can raise a page's visibility by up to 40%. We present the first systematic replication of this experiment in Brazilian Portuguese. Using 200 real queries in finance, health, and tourism, a simulated two-stage generative engine, and the nine original optimization methods with three repetitions per condition (N = 6,000 runs), we replicate the core pattern: statistics addition (+34.2% PAWC), quotation addition (+26.0%), and cite sources (+22.3%) yield significant gains, while keyword stuffing decreases visibility by 25.2%. We also observe the equalizer effect, with lower-ranked pages gaining proportionally more citations. Effects are domain-heterogeneous, with health being most conservative. GEO principles transfer to the Portuguese-language ecosystem.

**Keywords:** generative engine optimization; GEO; generative search; large language models; content visibility; Brazilian Portuguese.

## 1. Introdução

A forma dominante de acesso à informação na web está em transição. Sistemas como ChatGPT Search, Perplexity, Google AI Overviews e Gemini não devolvem mais uma lista de dez links azuis: eles recuperam um conjunto de páginas, sintetizam-nas em uma resposta única e atribuem trechos a fontes por meio de citações. Aggarwal et al. (2024) formalizaram esses sistemas como motores generativos (generative engines) e mostraram que essa mudança de paradigma redefine o que significa "estar visível" na web — a visibilidade deixa de ser uma posição em um ranking e passa a ser a fração da resposta gerada que é atribuída à página do criador de conteúdo.

No mesmo trabalho, os autores introduziram a Otimização para Motores Generativos (GEO): um conjunto de transformações de conteúdo — como adicionar estatísticas, citações diretas e referências a fontes — capazes de aumentar em 30–40% a visibilidade de uma página nas respostas geradas, e demonstraram que táticas clássicas de SEO, como keyword stuffing, são inócuas ou prejudiciais nesse novo regime. O estudo, conduzido sobre o benchmark GEO-bench com consultas em inglês, tornou-se o documento fundador de uma disciplina que cresce rapidamente na indústria e na academia.

Permanece, contudo, uma lacuna central: não se sabe se esses efeitos se transferem para outras línguas e ecossistemas informacionais. O português brasileiro é a quinta língua mais falada do mundo e o Brasil é um dos maiores mercados de busca e de adoção de assistentes de IA, mas a web lusófona difere da anglófona em densidade de fontes, convenções editoriais e confiabilidade percebida dos domínios. Este trabalho responde a essa lacuna com três contribuições: (i) a primeira replicação controlada do experimento GEO em português brasileiro, com 200 consultas reais em três domínios de alto impacto (finanças, saúde e turismo) e 6.000 execuções; (ii) evidência de que o padrão central do estudo original — superioridade dos métodos de autoridade epistêmica e fracasso do keyword stuffing — se reproduz em português, incluindo o efeito equalizador em favor de páginas mal ranqueadas; e (iii) a caracterização de heterogeneidades por domínio e por ancoragem local do conteúdo, com implicações práticas para criadores de conteúdo brasileiros.

## 2. Trabalhos relacionados

O estudo fundador de GEO (Aggarwal et al., 2024) definiu formalmente motores generativos, propôs métricas de visibilidade baseadas em contagem de palavras atribuídas e posição das citações e avaliou nove métodos de otimização sobre cerca de 10 mil consultas do GEO-bench, reportando ganhos de até 40% para métodos como adição de citações diretas e estatísticas, além do efeito equalizador — ganhos proporcionais maiores para fontes em posições baixas do ranking de recuperação. Trabalhos subsequentes ampliaram a agenda: Kumar e Lakkaraju (2024) mostraram que sequências adversariais podem manipular recomendações de produtos em LLMs; estudos de 2025–2026 propuseram bancadas específicas de e-commerce (E-GEO), otimização multiobjetivo de características de conteúdo e sistemas agênticos de GEO, e levantamentos críticos consolidaram a área. A literatura, entretanto, permanece quase exclusivamente anglófona; não há, até onde sabemos, replicação controlada dos métodos de Princeton em português, o que motiva este estudo.

# 3. Metodologia

## 3.1 Desenho experimental

Seguimos o protocolo de Aggarwal et al. (2024) com adaptações para o contexto brasileiro. Para cada consulta, um motor generativo simulado de duas etapas foi executado: (1) recuperação das cinco primeiras páginas orgânicas do Google para a consulta em português; (2) síntese, por um LLM, de uma resposta com citações no formato [n], recebendo as cinco fontes em ordem embaralhada de apresentação (slot 1–5). Uma das cinco páginas foi designada página-alvo, com sua posição orgânica original no Google (1–5) registrada na variável posicao_google e sua posição de apresentação no prompt registrada em slot_alvo.

## 3.2 Consultas e domínios

O banco contém 200 consultas reais em português brasileiro distribuídas em três domínios de alta relevância informacional e econômica: finanças (71 consultas), turismo (71) e saúde (58). Registramos também a ancoragem brasileira do conteúdo-alvo (ancoragem_br $\in \{0,1\}$), indicando se a página está fortemente ancorada em entidades, instituições e dados brasileiros (102 consultas) ou trata de tema internacional/genérico (98), permitindo testar se sinais de localidade moderam os efeitos de otimização.

## 3.3 Métodos de otimização

Aplicamos à página-alvo os nove métodos do estudo original, traduzidos e adaptados ao português: citação de fontes (cite sources), adição de citações diretas (quotation addition), adição de estatísticas (statistics addition), tom autoritativo (authoritative), otimização de fluência (fluency optimization), linguagem acessível (easy-to-understand), termos técnicos (technical terms), palavras únicas (unique words) e keyword stuffing, além da condição baseline sem alteração. As transformações preservam o conteúdo factual da página (alterando em média +53 a +95 palavras; variável delta_n_palavras) e cada versão passou por uma checagem automática de fidelidade da manipulação (check_manipulacao), com taxas de aprovação entre 84% e 90%.

## 3.4 Métricas e análise estatística

Adotamos as métricas de impressão do estudo original: PAWC (Position-Adjusted Word Count), fração de palavras da resposta atribuída à página-alvo ponderada pela posição das sentenças que a citam; a fração bruta de palavras atribuídas (wordcount_sub); o escore posicional (position_sub); e a taxa de citação (citou_alvo), indicando se a página-alvo foi citada ao menos uma vez. Cada condição consulta × método foi executada 3 vezes (temperatura não nula), totalizando 200 × 10 × 3 = 6.000 execuções. Respostas em que o modelo se recusou a responder (2,2% em média, sobretudo em saúde) foram registradas na variável recusou. Para inferência, agregamos as repetições por consulta e comparamos cada método ao baseline por teste de Wilcoxon pareado por consulta (nível de significância de 5%), reportando a melhoria relativa percentual das médias, como no estudo original.

# 4. Resultados

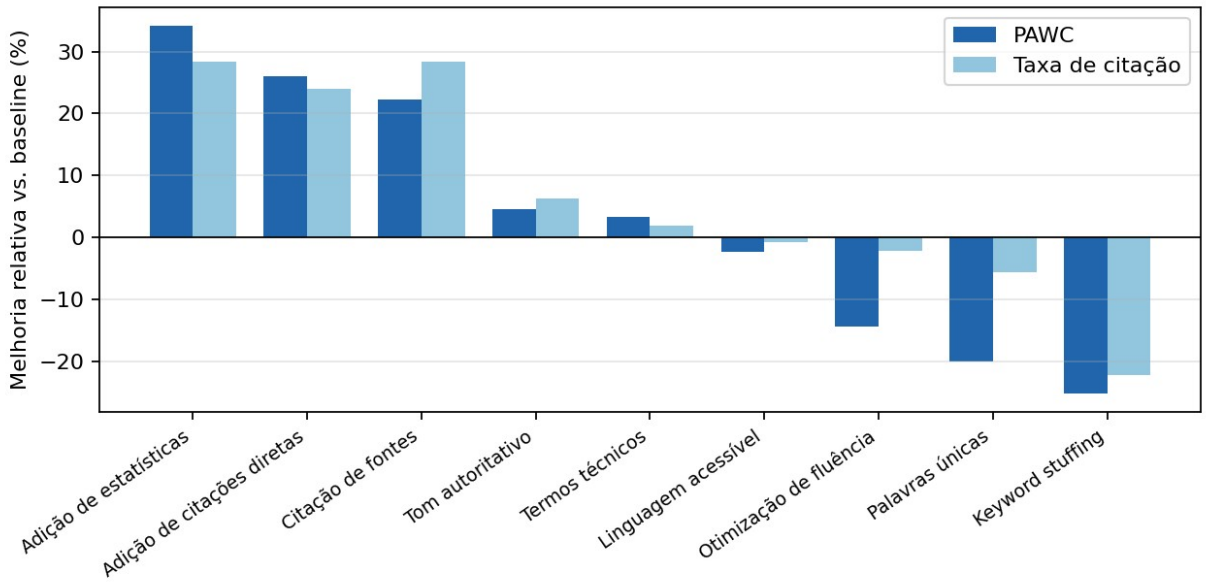
## 4.1 Efeito geral dos métodos

A Tabela 1 e a Figura 1 apresentam o efeito de cada método sobre as duas métricas principais. O padrão central do estudo de Princeton se replica integralmente em português: os três métodos de autoridade epistêmica — adição de estatísticas, adição de citações diretas e citação de fontes — dominam o ranking, com ganhos de PAWC de +22% a +34% e ganhos de taxa de citação de +24% a +28%, todos estatisticamente significativos ($p < 0,01$). No extremo oposto, keyword stuffing, a tática canônica do SEO tradicional, reduz o PAWC em 25,2% e a probabilidade de citação em 22,1% ($p < 0,01$) — replicando e até ampliando o achado original de que sinais clássicos de densidade de palavras-chave são contraproducentes em motores generativos. Métodos puramente estilísticos (tom autoritativo, termos técnicos, linguagem acessível) têm efeitos nulos, e a otimização de fluência e a substituição por palavras únicas tendem a reduzir a visibilidade.

Tabela 1 — Visibilidade média da página-alvo por método (médias por consulta; N = 200 consultas, 3 repetições). Δ = melhoria relativa vs. baseline; p do teste de Wilcoxon pareado sobre PAWC; n.s. = não significativo.

| Método | PAWC | Δ PAWC | Taxa de citação | Δ Citação | p (Wilcoxon, PAWC) |
|---|---|---|---|---|---|
| Baseline (sem otimização) | 0,104 | — | 45,2% | — | — |
| Adição de estatísticas | 0,139 | +34,2% | 58,0% | +28,4% | < 0,001 |
| Adição de citações diretas | 0,131 | +26,0% | 56,0% | +24,0% | < 0,001 |
| Citação de fontes | 0,127 | +22,3% | 58,0% | +28,4% | 0,005 |
| Tom autoritativo | 0,108 | +4,6% | 48,0% | +6,3% | 0,831 (n.s.) |
| Termos técnicos | 0,107 | +3,3% | 46,0% | +1,8% | 0,959 (n.s.) |
| Linguagem acessível | 0,101 | −2,3% | 44,8% | −0,7% | 0,950 (n.s.) |
| Otimização de fluência | 0,089 | −14,4% | 44,2% | −2,2% | 0,051 (n.s.) |
| Palavras únicas | 0,083 | −20,0% | 42,7% | −5,5% | 0,027 |
| Keyword stuffing | 0,078 | −25,2% | 35,2% | −22,1% | 0,005 |

Figura 1 — Impacto dos métodos GEO na visibilidade (200 consultas × 3 repetições)

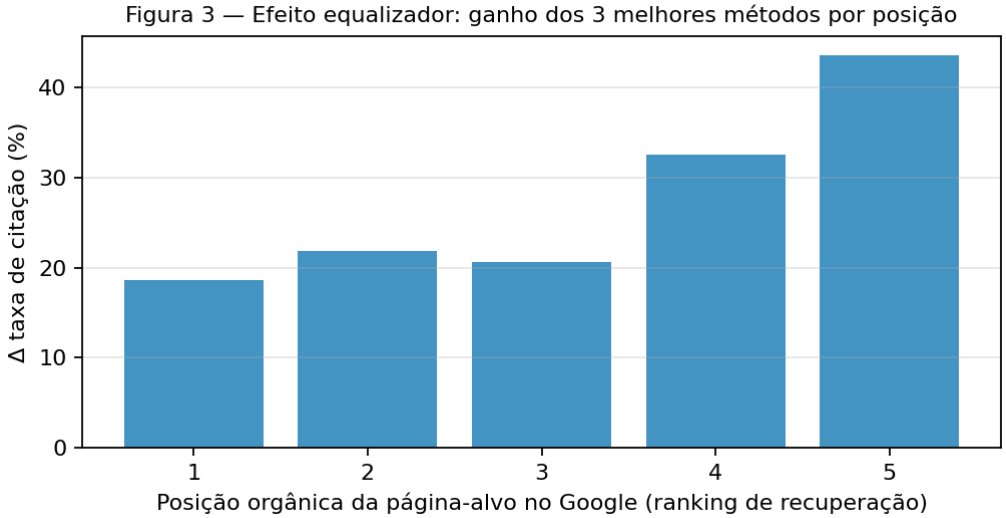



*Fonte: dados do experimento. Barras ordenadas pelo ganho em PAWC.*

## 4.2 Efeito equalizador

O estudo original reportou que a otimização beneficia desproporcionalmente fontes mal posicionadas no ranking de recuperação. Reproduzimos o teste agrupando os três melhores métodos e estratificando pelo ranking orgânico da página-alvo no Google (Figura 3). O ganho relativo na taxa de citação cresce monotonicamente da posição 1 (+18,6%) para as posições 4 (+32,6%) e 5 (+43,6%). Em outras palavras, no ecossistema brasileiro os motores generativos também funcionam como um equalizador: páginas sem tráfego de topo de ranking são justamente as que mais têm a ganhar com GEO, um resultado alinhado ao efeito descrito por Aggarwal et al. (2024) — ainda que com magnitude menor que os +115% reportados para a posição 5 no corpus em inglês. Observamos ainda um forte efeito da posição de apresentação no prompt: quando a página-alvo aparece no último slot do contexto, sua taxa de citação no baseline cai de 53% (slot 1) para 31% (slot 5), evidenciando viés posicional do LLM sintetizador.

Figura 3 — Efeito equalizador: ganho dos 3 melhores métodos por posição

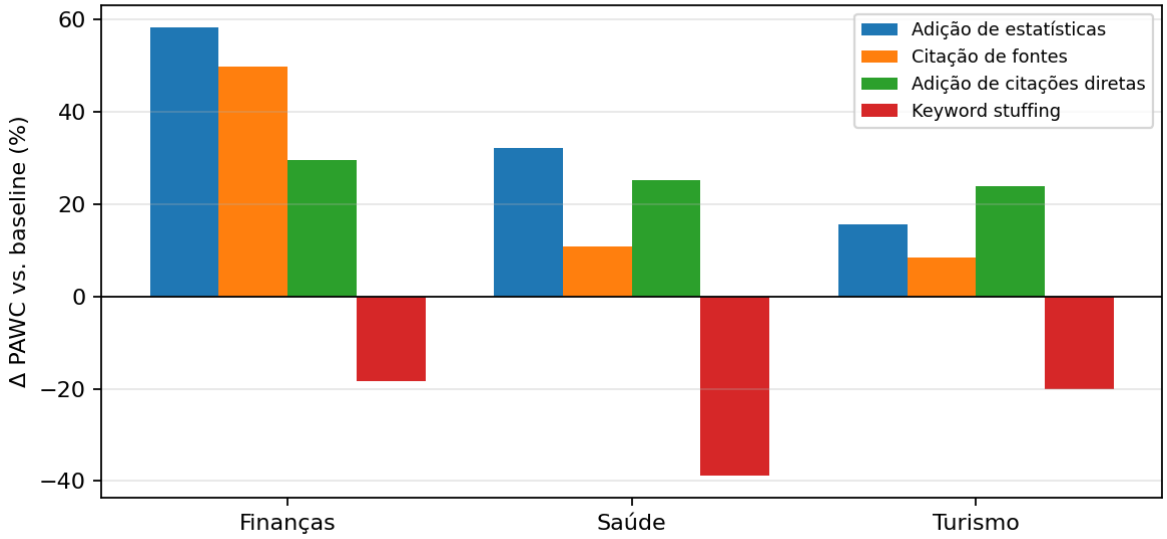



*Fonte: dados do experimento. Métodos agrupados: estatísticas, citações diretas e fontes.*

### 4.3 Heterogeneidade por domínio e ancoragem local

A Figura 2 mostra que a eficácia dos métodos varia substancialmente entre domínios. Em finanças, a adição de estatísticas atinge +58,3% e a citação de fontes +49,8% — os maiores efeitos do estudo —, coerente com consultas cuja resposta ideal é intrinsecamente quantitativa. Em turismo, os ganhos são positivos porém moderados (+8% a +24%). Saúde é o domínio mais conservador: apenas os métodos de evidência direta (estatísticas, +32,2%; citações diretas, +25,2%) produzem ganhos claros, enquanto tom autoritativo (−20,1%) e otimização de fluência (−39,8%) chegam a ser penalizados, sugerindo que o sintetizador privilegia fontes com evidência verificável — e não com retórica de autoridade — em temas sensíveis; é também o domínio com maior taxa de recusa do modelo. Quanto à ancoragem local, os métodos vencedores mantêm eficácia em ambos os estratos, mas a adição de citações diretas é mais eficaz em conteúdo internacional (+42,0%) do que em conteúdo fortemente ancorado no Brasil (+10,8%), enquanto a adição de estatísticas é robusta nos dois casos (+34,2% em ambos) — indicando que estatísticas são o sinal de autoridade mais transferível entre contextos.

Figura 2 — Heterogeneidade dos efeitos por domínio




*Fonte: dados do experimento. Δ PAWC relativo ao baseline de cada domínio.*

## 5. Discussão

Os resultados sustentam três conclusões. Primeiro, os princípios de GEO se transferem para o português brasileiro: a hierarquia de métodos observada — evidência quantitativa e atribuição de fontes no topo, manipulação lexical no fundo — é essencialmente a mesma do corpus em inglês, com magnitudes comparáveis (nossos +34% para estatísticas versus os +41% de Princeton na mesma métrica). Isso sugere que o comportamento de citação dos LLMs sintetizadores reflete preferências profundas por autoridade epistêmica, e não idiossincrasias do inglês. Segundo, a morte do keyword stuffing confirma-se de forma ainda mais acentuada: no regime generativo, a tática não é apenas inútil, é ativamente punida — um alerta direto para o mercado brasileiro de SEO, que ainda investe pesadamente em densidade de palavras-chave. Terceiro, o efeito equalizador tem implicação distributiva relevante para a web lusófona, cujo mercado de busca é altamente concentrado: produtores de conteúdo pequenos e mal ranqueados são os maiores

beneficiários potenciais de GEO, o que pode reduzir — ou, se apenas grandes players adotarem GEO, ampliar — assimetrias de visibilidade.

A heterogeneidade em saúde merece destaque normativo. O fato de o sintetizador recompensar evidência verificável e punir retórica autoritativa em temas médicos é, em princípio, uma propriedade desejável; mas também mostra que técnicas de GEO poderiam ser exploradas para dar proeminência a conteúdo de saúde de baixa qualidade que simplesmente adote a forma da evidência (estatísticas e citações fabricadas). Isso reforça a necessidade, já apontada na literatura, de mecanismos de verificação de fontes nos motores generativos.

## 6. Limitações

Cinco limitações delimitam o alcance dos achados. (1) O motor generativo é simulado com um único LLM sintetizador; motores comerciais combinam recuperação proprietária, reranqueamento e modelos distintos, e os efeitos podem variar entre plataformas. (2) O banco de 200 consultas em três domínios é menor que o GEO-bench (~10 mil consultas, 9 conjuntos), limitando o poder para efeitos pequenos e a generalização a outros domínios. (3) As transformações de conteúdo foram geradas automaticamente e, embora auditadas (84–90% de aprovação na checagem de fidelidade), podem introduzir variações não intencionais. (4) Durante a preparação dos dados identificamos e corrigimos uma corrupção de localidade decimal nas colunas de métricas contínuas (pontos decimais interpretados como separador de milhar); a recuperação foi exata para a fração de palavras (validada por reconstrução fracionária k/n) e por consistência de magnitude para PAWC e escore posicional, procedimento documentado no material suplementar. (5) Medimos visibilidade na resposta, não cliques ou conversões; a relação entre citação em motores generativos e tráfego real permanece questão aberta.

## 7. Conclusão

Este trabalho apresentou a primeira replicação controlada, em português brasileiro, do experimento fundador da Otimização para Motores Generativos. Com 6.000 execuções sobre 200 consultas reais de finanças, saúde e turismo, mostramos que os achados centrais de Aggarwal et al. (2024) se reproduzem no ecossistema lusófono: adicionar estatísticas, citações diretas e referências a fontes eleva a visibilidade em 22–34%, keyword stuffing a reduz em 25%, e páginas mal ranqueadas são as que proporcionalmente mais ganham. Para criadores de conteúdo brasileiros, a diretriz empírica é clara: no regime da busca generativa, autoridade epistêmica demonstrável vale mais do que qualquer manipulação lexical. Trabalhos futuros devem ampliar o banco de consultas, testar motores comerciais reais (AI Overviews, Perplexity, ChatGPT Search) em português, medir efeitos longitudinais sobre tráfego e investigar defesas contra o uso adversarial das próprias técnicas aqui validadas.

## Referências

AGGARWAL, P.; MURAHARI, V.; RAJPUROHIT, T.; KALYAN, A.; NARASIMHAN, K.; DESHPANDE, A. GEO: Generative Engine Optimization. In: PROCEEDINGS OF THE 30TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (KDD ’24), Barcelona, 2024. New York: ACM, 2024. p. 5–16. DOI: 10.1145/3637528.3671900.

KUMAR, A.; LAKKARAJU, H. Manipulating Large Language Models to Increase Product Visibility. arXiv:2404.07981, 2024.

BAGGA, P. S.; FARIAS, V. F.; KORKOTASHVILI, T.; PENG, T.; WU, Y. E-GEO: A Testbed for Generative Engine Optimization in E-commerce. arXiv:2511.20867, 2025.

LIU, Z.; XU, P. Think Before Writing: Feature-Level Multi-Objective Optimization for Generative Citation Visibility. arXiv:2604.19113, 2026.

WAN, A.; WALLACE, E.; KLEIN, D. What Evidence Do Language Models Find Convincing? In: PROCEEDINGS OF THE 62ND ANNUAL MEETING OF THE ACL. Stroudsburg: ACL, 2024. p. 7468–7484.